

outlines

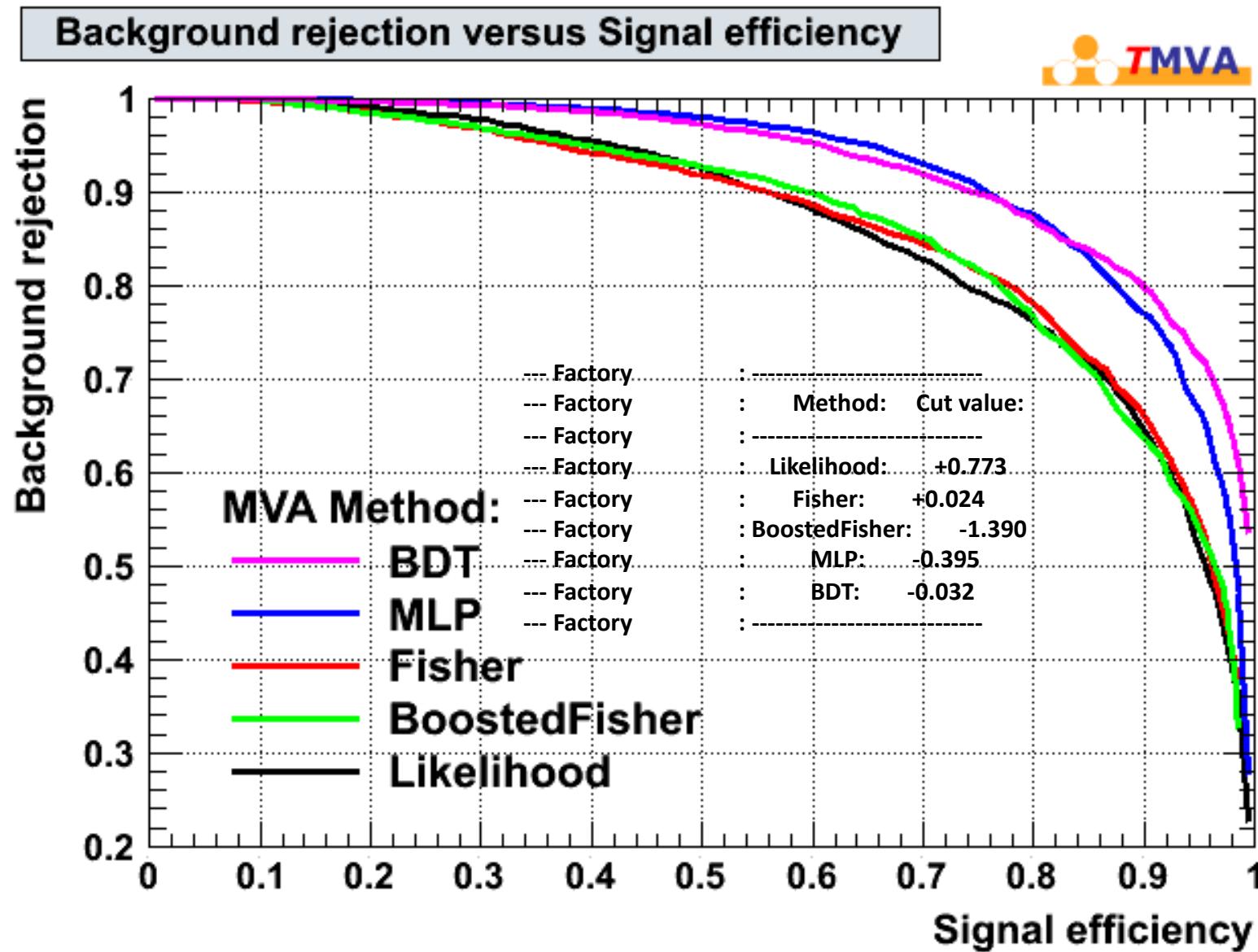
- Signal : single D⁰ power law : ~43k
 - After cuts to match the 3rd production → ~9k
- Background : unlike sign pairs from real data
 - files from day 113, day 120, day 134 → check the stability and if we can use only 1 file to process the whole run 7.
 - After cut ($Dca3D^* < 1$) → ~39,40k (see slide 18)
 - It is recommended to have large and equal (in size) background and signal samples
 - Here the background sample is x3 the signal sample
 - Not an issue (the opposite would have been)
 - I've tested with similar samples size and did not see dramatic changes in the mva responses.
- Classifiers : Likelihood, BDT, Fisher, MLP, BoostedFisher

* Dca3D : is the distance between 2 tracks at secondary vertex point

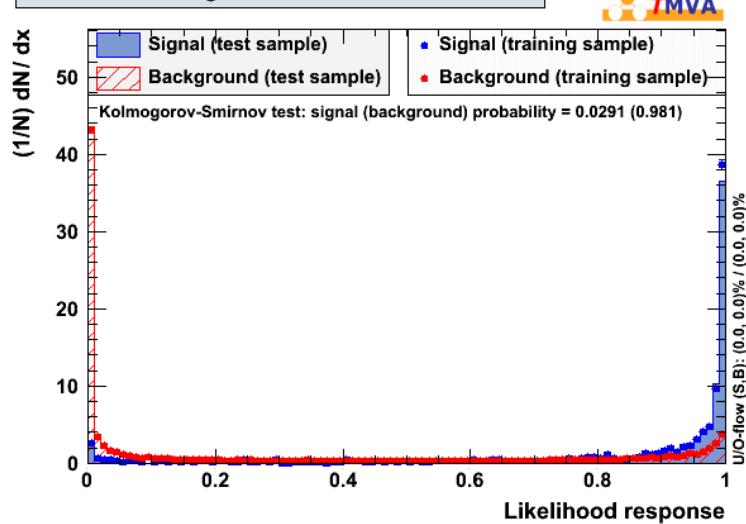
Real data day 113 trained with background from day 113

BACKGROUND : DAY 113

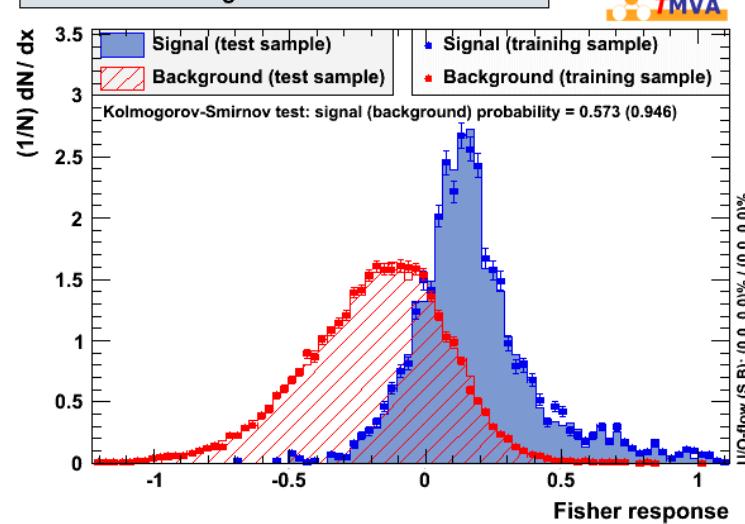
Signal eff. Vs. background rej.



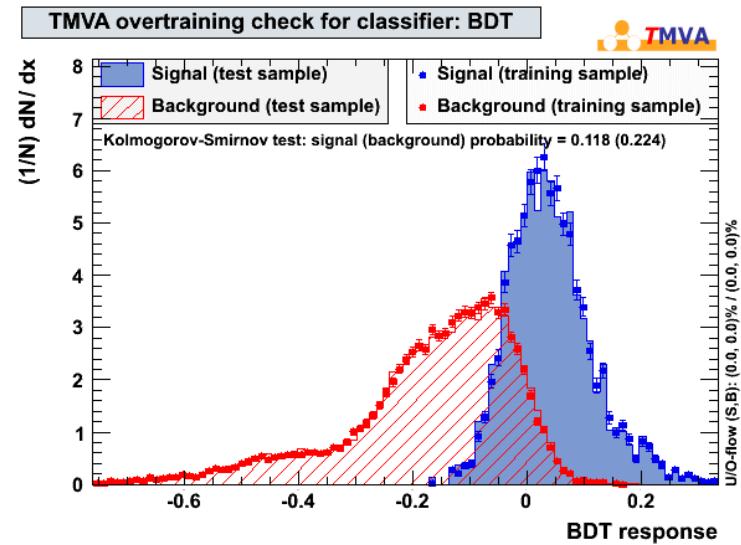
TMVA overtraining check for classifier: Likelihood



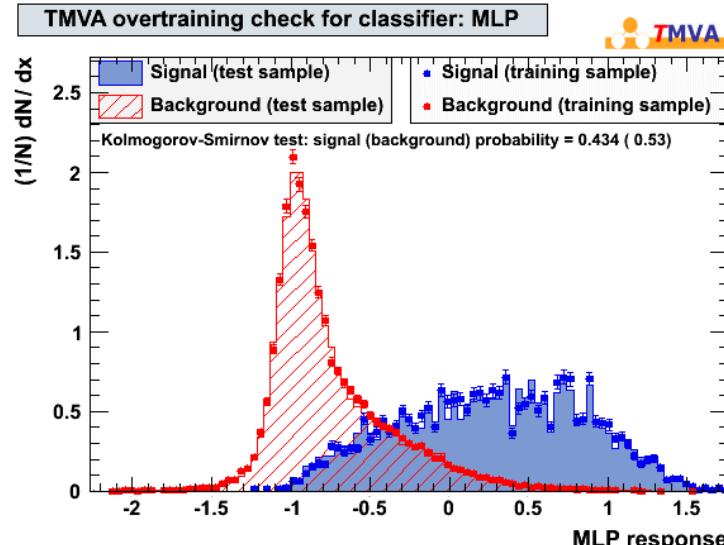
TMVA overtraining check for classifier: Fisher



TMVA overtraining check for classifier: BDT

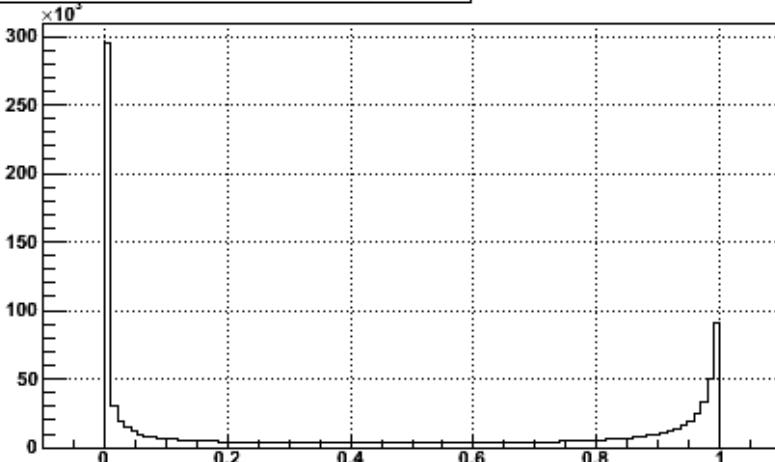


TMVA overtraining check for classifier: MLP

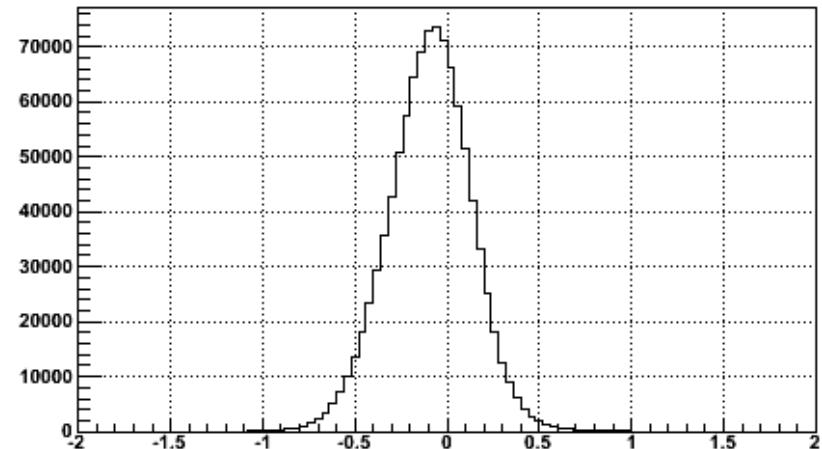


Mva responses distribution after rerunning real data

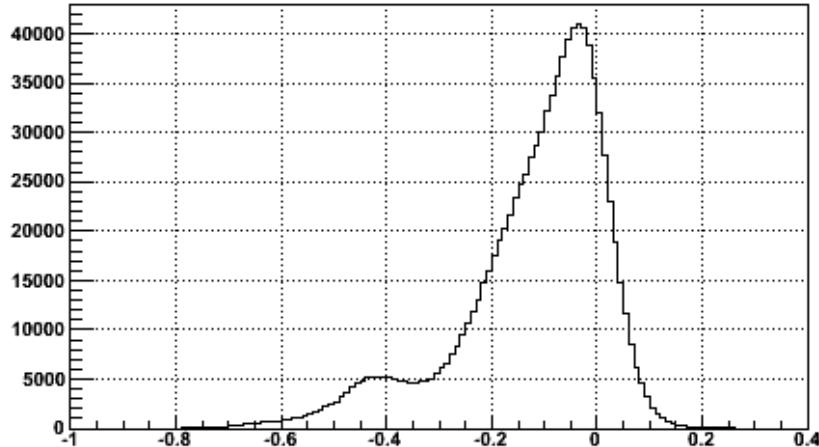
mva Likelihood response for background



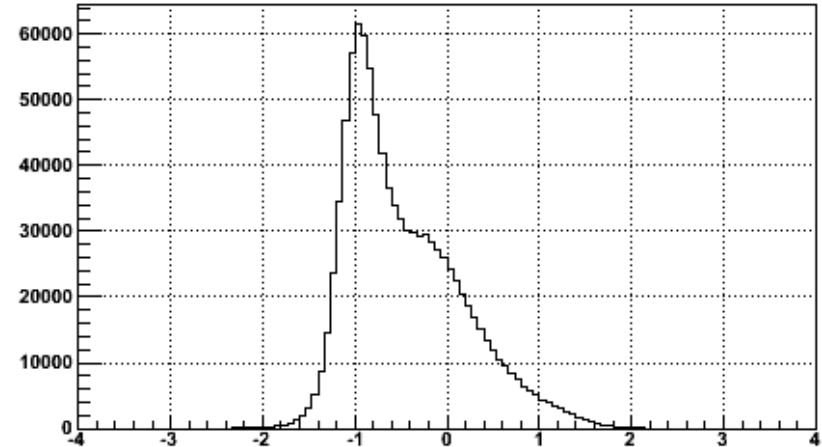
mva FISHER response for background



mva BDT response for background



mva MLP response for background



- it has to match the red curves (previous slide)

comments

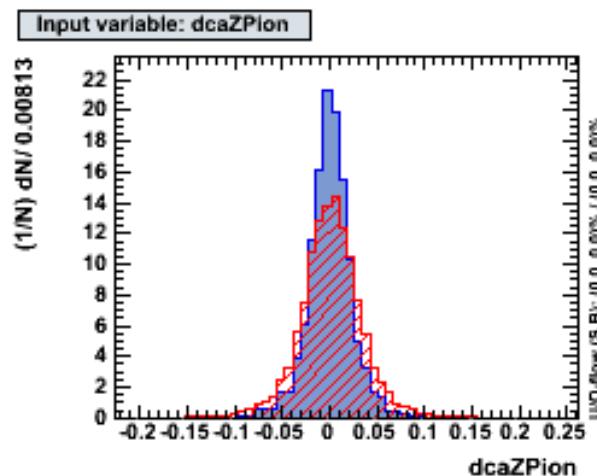
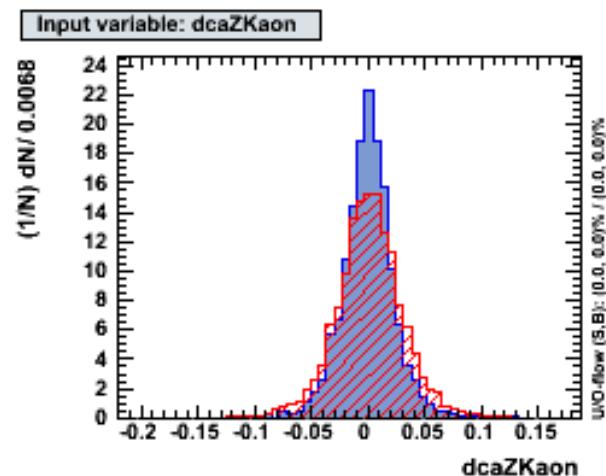
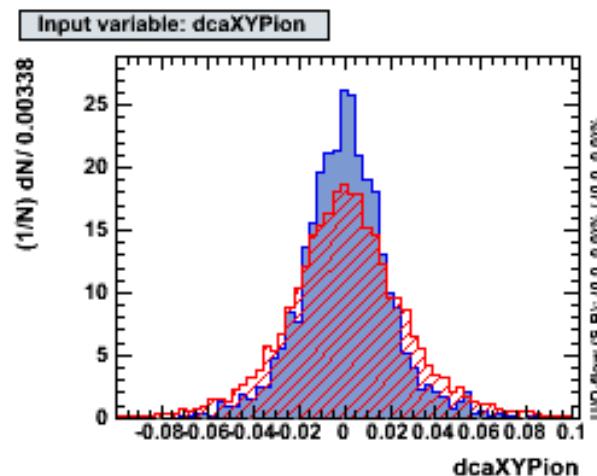
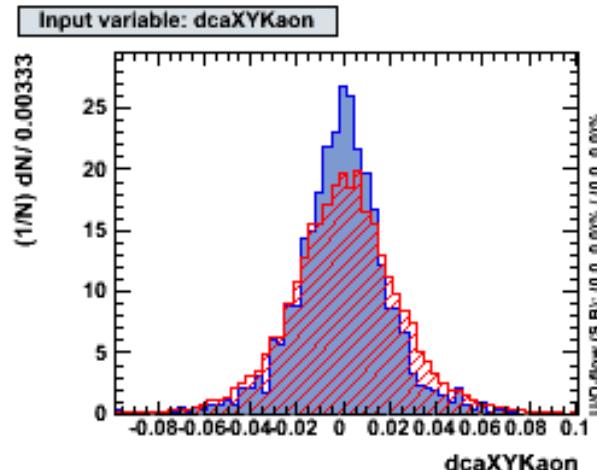
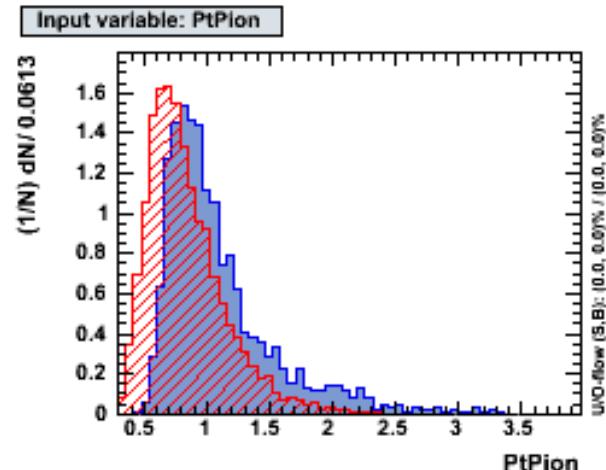
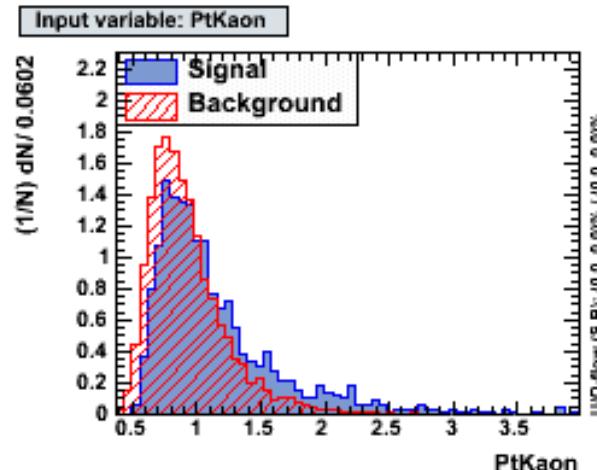
- For mvaMLP : same peak ~ -1
- For BDT : same peaks ~ -0.4 and ~ -0.05 (the second is less pronounced)
- For Fisher : peak ~ -0.1

Summary : best cut according TMVA

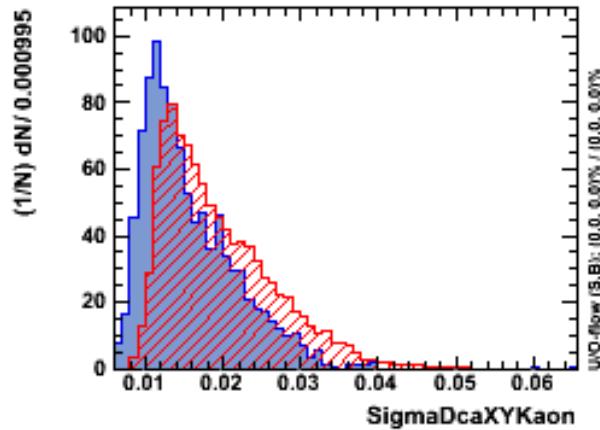
MVA	#events (background)	MLP	Fisher	BDT	Likelihood	Boosted Fisher
Day 113	33908	-0.395	0.024	-0.032	0.773	-1.390
Day 120	33908	-0.369	0.023	-0.041	0.765	-1.361
Day 134	33823	-0.305	0.025	-0.029	0.795	-1.395

- Given the range of each mva responses, the difference when taking background from different day is very small
- Stability of the classifier

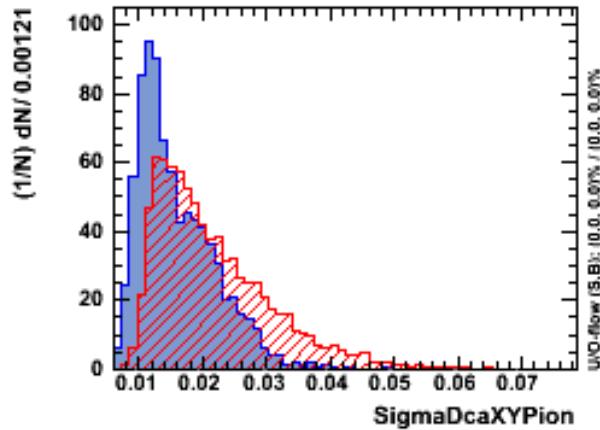
TRAINING SAMPLES : SIGNAL AND BACKGROUND VARIABLES DISTRIBUTIONS



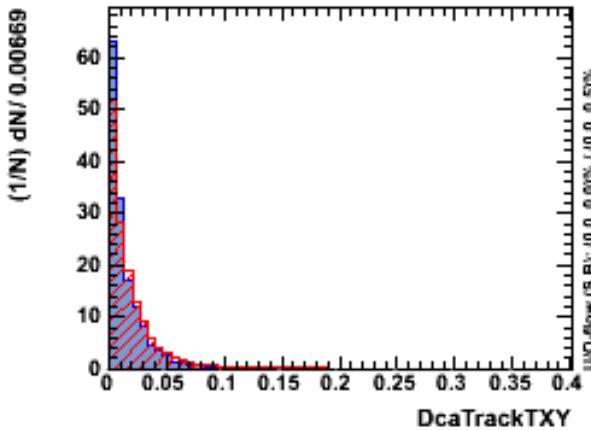
Input variable: SigmaDcaXYKaon



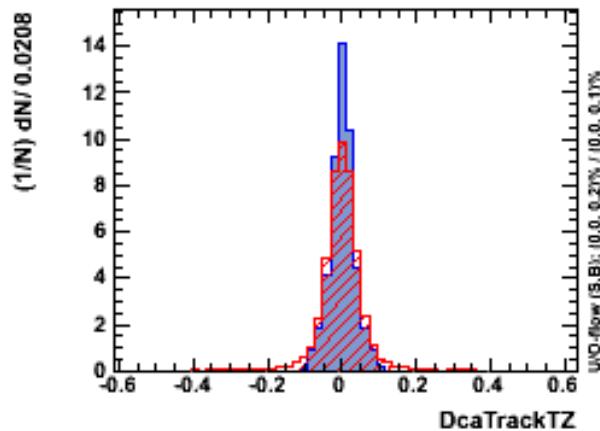
Input variable: SigmaDcaXYPion



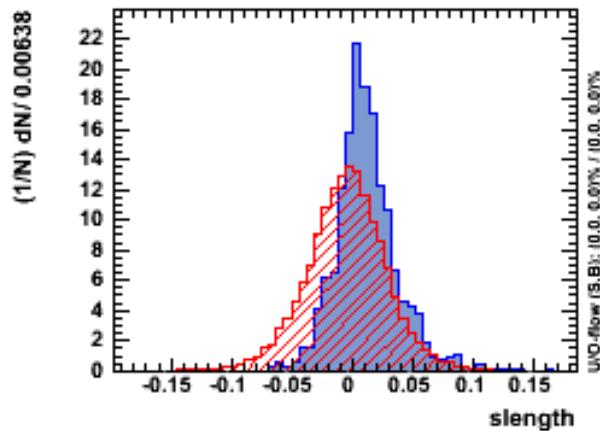
Input variable: DcaTrackTXY



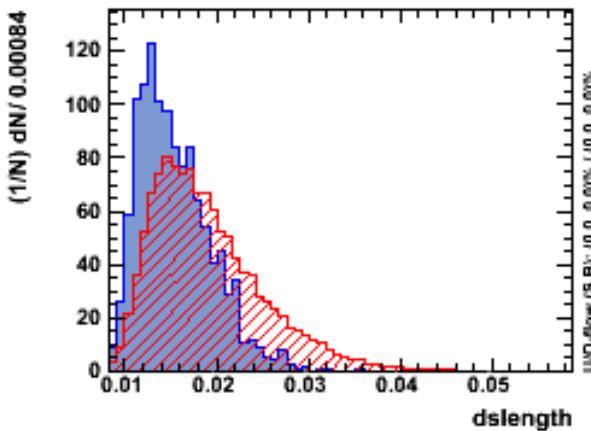
Input variable: DcaTrackTZ

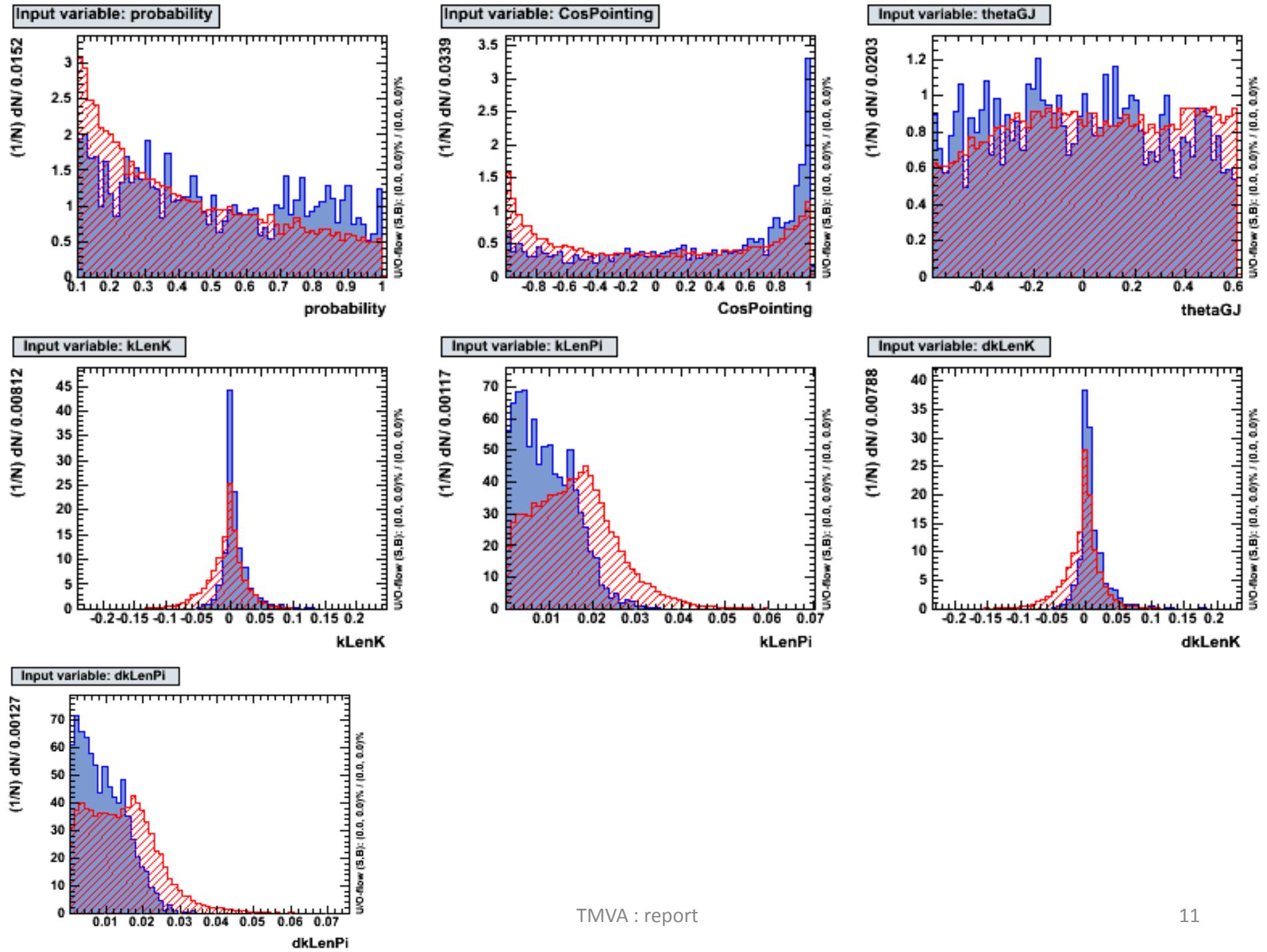


Input variable: slength



Input variable: dslength

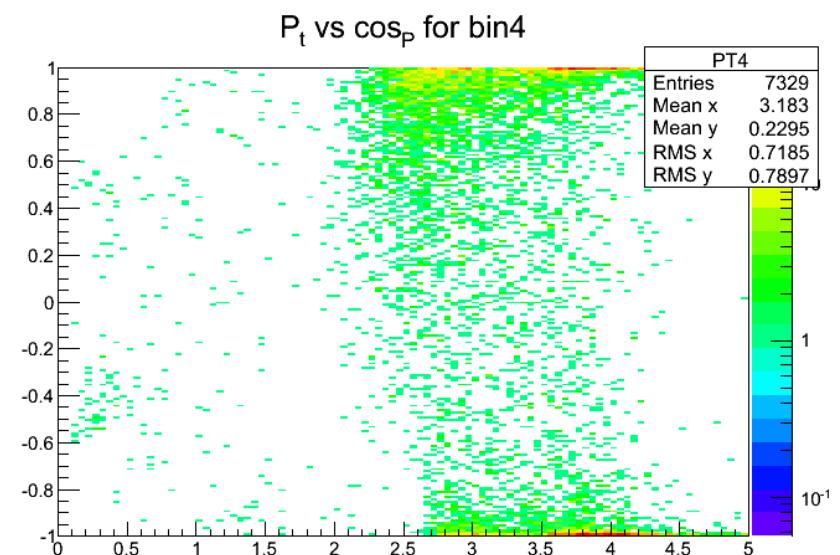
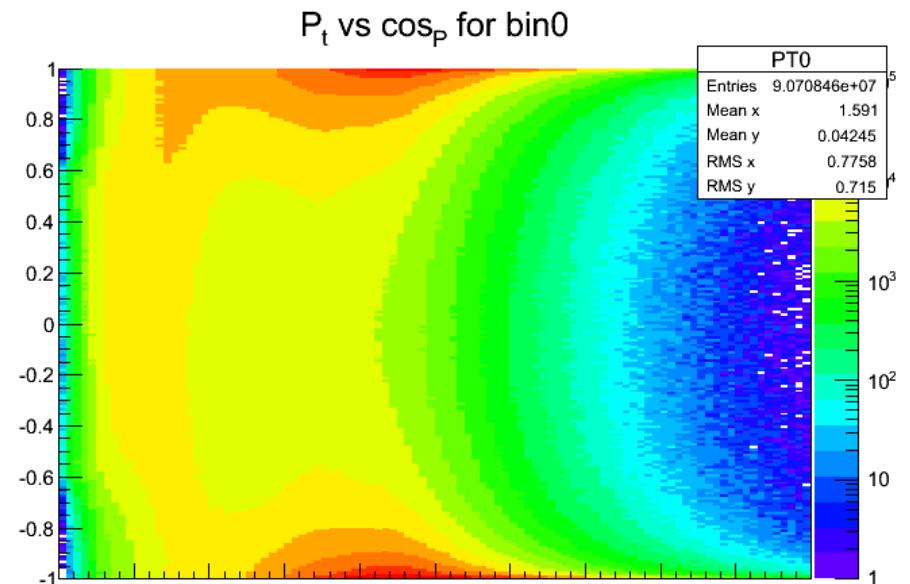
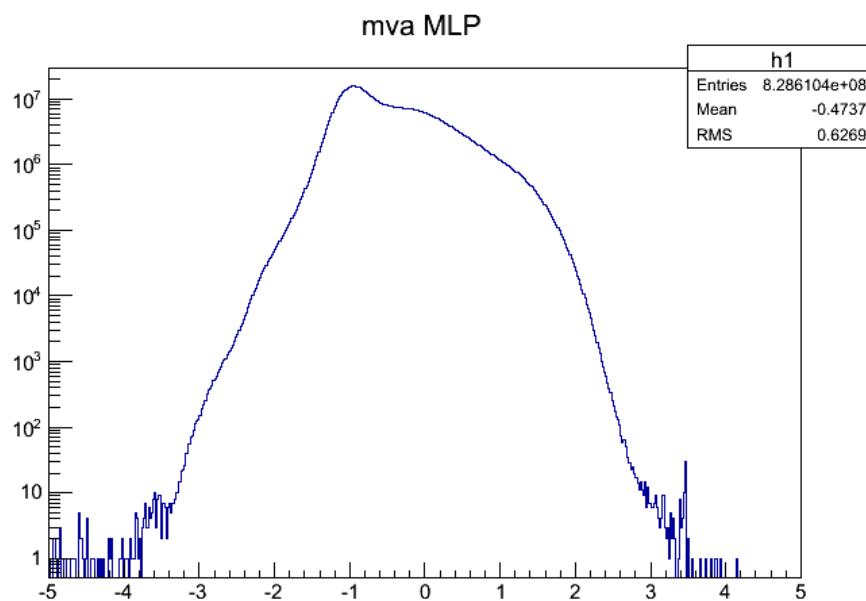




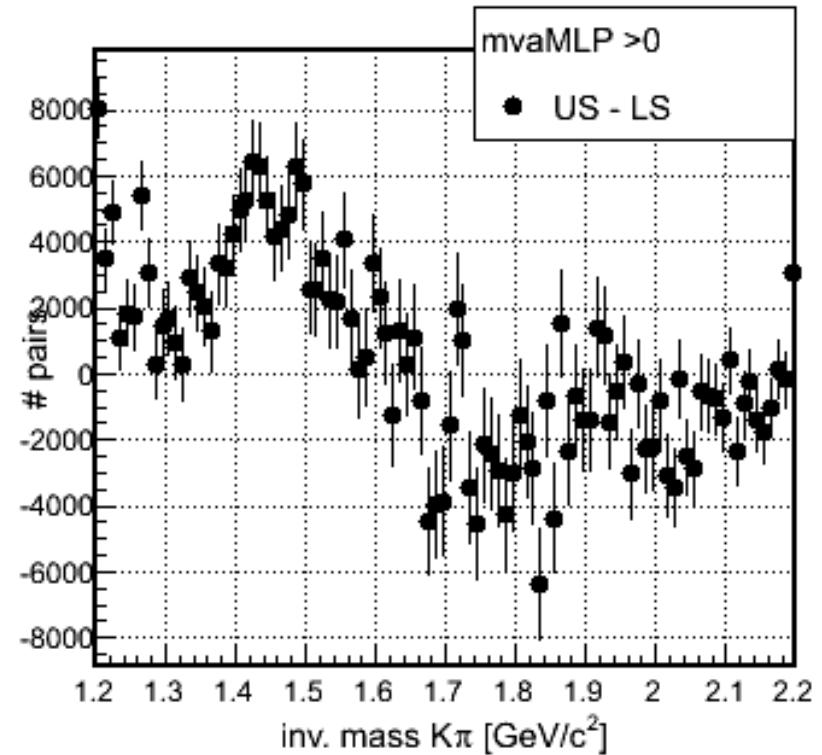
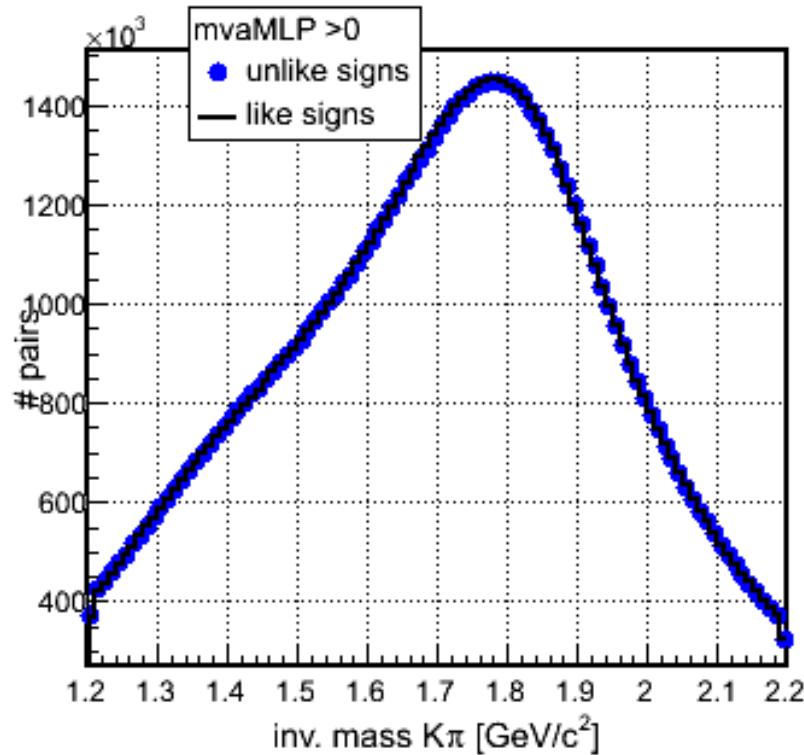
statistic

Files from Day XX to day YY	98 99	100 109	110 119	120 129	130 139	140 149	150 159	160 169	170 177
# of events									
# of pairs(before Dca3D cut)	13647200	30738112	186221907	152479979	106247554	128205573	44277293	107944064	67721308
# of pairs (after Dca3D cut)	13522000	30434788	184467807	150998510	105180970	126724558	43745994	106644134	66891680

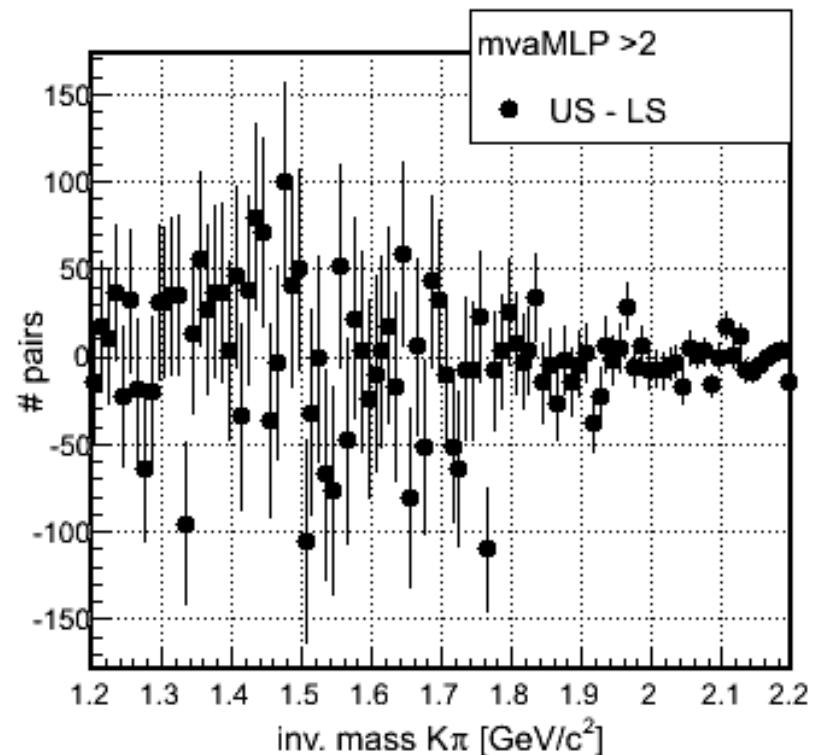
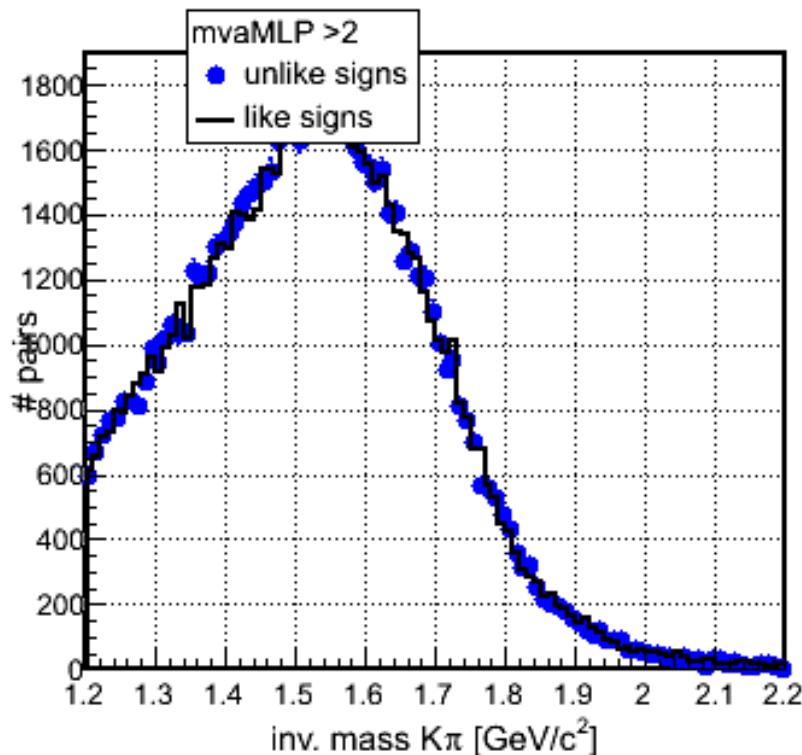
Results for mvaMLP



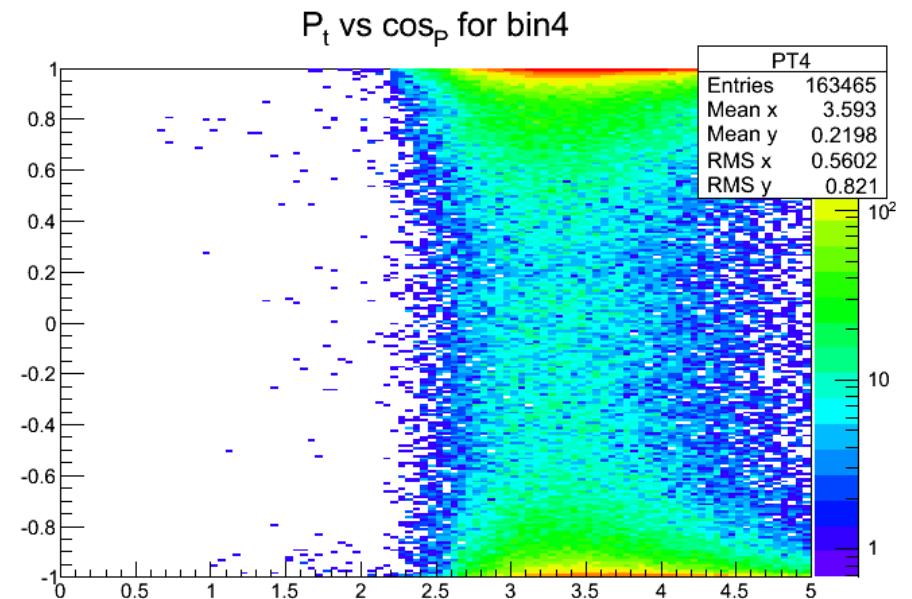
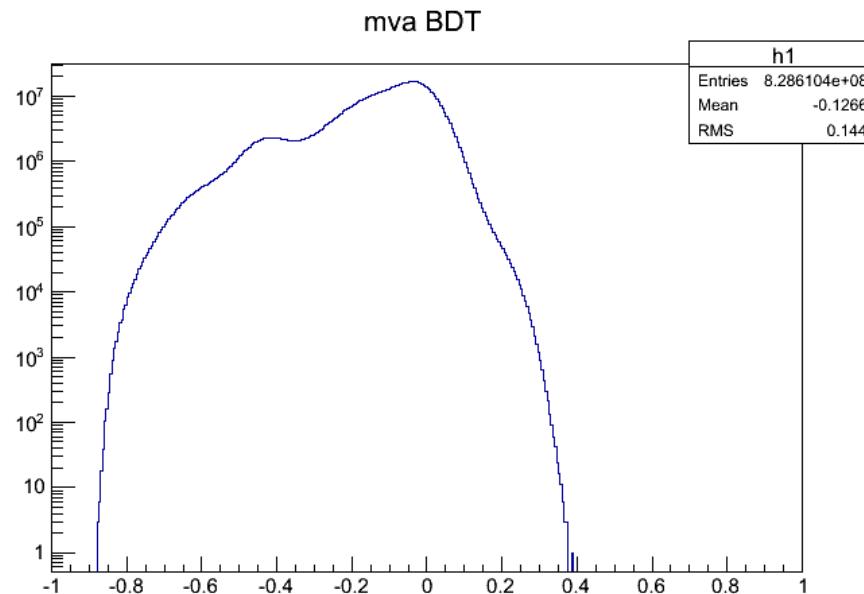
Invariant mass (mvaMLP>0)



Invariant mass (mvaMLP>2)

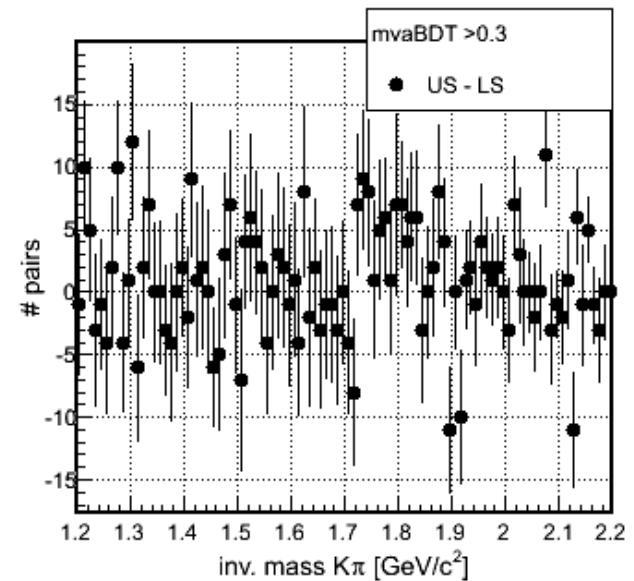
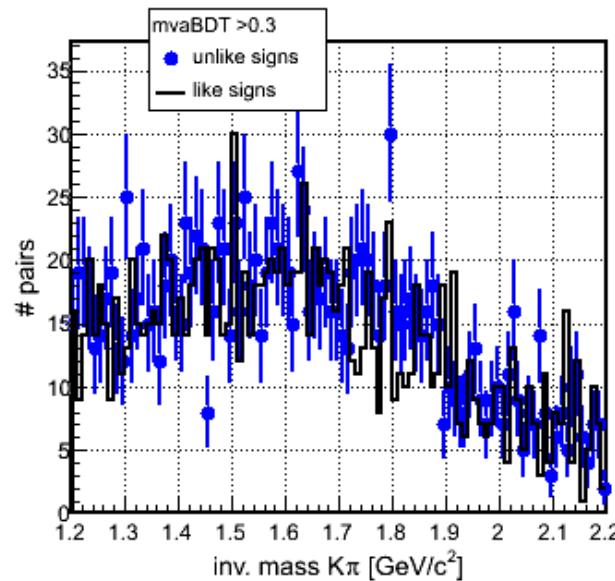
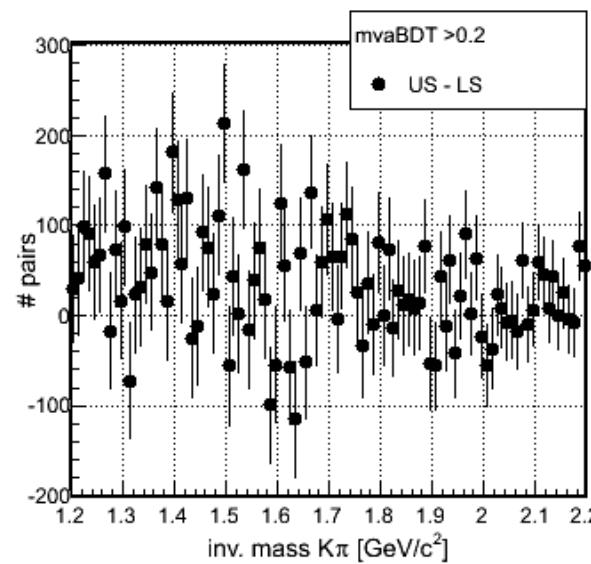
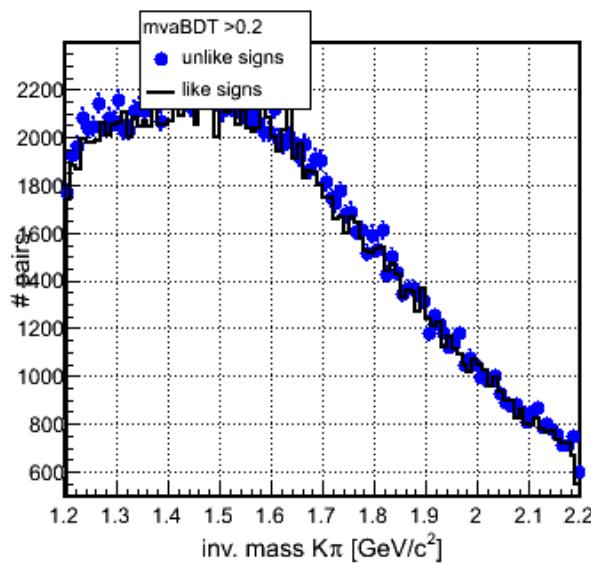


Results for mvaBDT

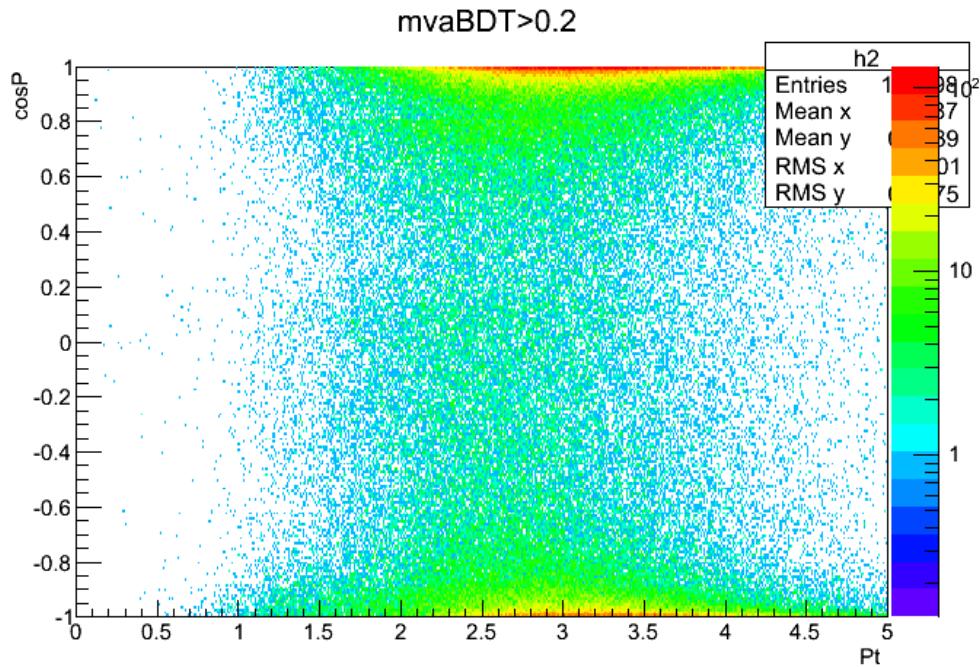


- BIN4 is for mvaBDT>0.2
- Same observation than for mvaMLP : higher mva responses give “high” Pt pairs

Invariant mass



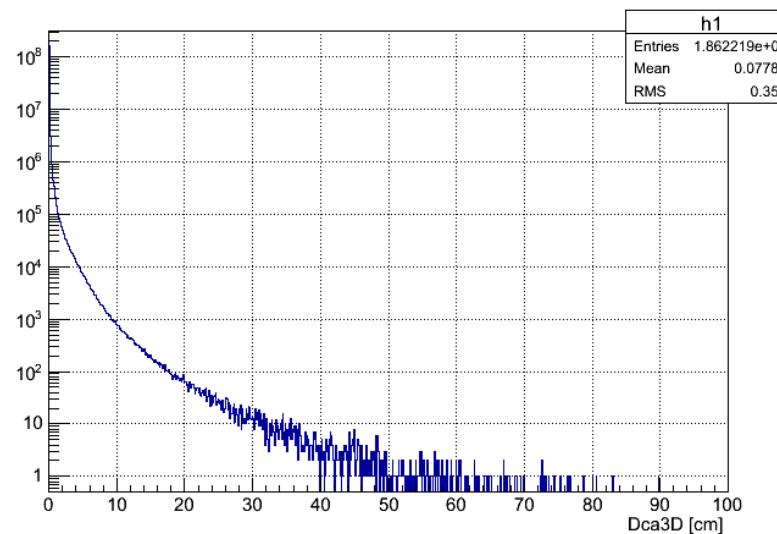
Next steps



- I'm rerunning with a training done w/o the P_t of daughters
- Left plot : for the same mva response, it looks that the selection of high P_t pairs is less (but is still here, I think because of the decay length)

Dca3D distribution

Before cut



After cut [1cm]

